AD NO. ___________________________
ATEC PROJECT NO. 2011-DT-ATC-DODSP-F0292
REPORT NO. ATC 11583

# STANDARDIZED

# UXO TECHNOLOGY DEMONSTRATION SITE

# SCORING RECORD NO. 943

SITE LOCATION:
ABERDEEN PROVING GROUND

DEMONSTRATOR:
BATTELLE
100A DONNER DRIVE
OAK RIDGE, TN 37830

TECHNOLOGY TYPE/PLATFORM:
TEM-8G
TOWED ARRAY

AREAS COVERED:
SMALL MUNITIONS TEST SITE

PREPARED BY:
U.S. ARMY ABERDEEN TEST CENTER
ABERDEEN PROVING GROUND, MD 21005-5059

AUGUST 2014

Prepared for:
SERDP/ESTCP
MUNITIONS MANAGEMENT
ALEXANDRIA, VA 22350

U.S. ARMY TEST AND EVALUATION COMMAND
ABERDEEN PROVING GROUND, MD 21005-5001

## DISPOSITION INSTRUCTIONS

**DEPARTMENT OF THE ARMY**
**U.S. ARMY ABERDEEN TEST CENTER**
**400 COLLERAN ROAD**
**ABERDEEN PROVING GROUND, MARYLAND 21005-5059**

REPLY TO
ATTENTION OF

TEDT-AT-SL-M

MEMORANDUM FOR Program Manager – SERDP/ESTCP, Munitions Management, Mr. Herb Nelson, 4800 Mark Center Drive, Suite 17D08, Alexandria, VA 22350-3600

SUBJECT: Standardized Unexploded Ordinance (UXO) Technology Demonstration Site Scoring Record No. 943

1. The subject Scoring Record is submitted for your information and retention.

2. The point of contact for this office is Mr. Leonard Lombardo, Test Officer, Survivability/Lethality Directorate, Maritime/Threat Detection Systems Survivability Division (TDSS) and may be reached at 410-278-1286 or via e-mail, leonard.c.lombardo.civ@mail.mil.

TRUTH IN TESTING!

FOR THE COMMANDER:

SHEPPARD.TRACY.V.1200534219

Digitally signed by SHEPPARD.TRACY.V.1200534219
DN: c=US, o=U.S. Government, ou=DoD, ou=PKI, ou=USA, cn=SHEPPARD.TRACY.V.1200534219
Date: 2014.11.09 20:03:09 -05'00'

Encl

TRACY V. SHEPPARD
Director, Survivability/Lethality Directorate

| **REPORT DOCUMENTATION PAGE** | *Form Approved OMB No. 0704-0188* |
|---|---|



| **1. REPORT DATE** *(DD-MM-YYYY)* | **2. REPORT TYPE** | **3. DATES COVERED** *(From - To)* |
|---|---|---|
| August 2014 | Final | May 2013 |

**4. TITLE AND SUBTITLE**
Standardized UXO Technology Demonstration Site
Scoring Record No. 943

**5a. CONTRACT NUMBER**

**5b. GRANT NUMBER**

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**
Hill, Scott
Lombardo, Leonard
Fling Jr., Rick

**5d. PROJECT NUMBER**
2011-DT-ATC-DODSP-F0292

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
Commander
U.S. Army Aberdeen Test Center
ATTN: TEDT-AT-SLE
Aberdeen Proving Ground, MD 21005-5059

**8. PERFORMING ORGANIZATION REPORT NUMBER**
ATC-11583

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
Program Manager
SERDP/ESTCP
Munitions Management
Alexandria, VA 22350

**10. SPONSOR/MONITOR'S ACRONYM(S)**

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**
Same as Item 8

**12. DISTRIBUTION/AVAILABILITY STATEMENT**
Distribution unlimited.

**13. SUPPLEMENTARY NOTES**
None

**14. ABSTRACT**

This scoring record documents the efforts of the Environmental Security Technology Certification Program to detect and discriminate inert unexploded ordnance (UXO) utilizing the APG Standardized UXO Technology Demonstration Site calibration lanes and open field sites. This Scoring Record was coordinated by J. Stephen McClung and the Standardized UXO Technology Demonstration Site Scoring Committee. Organizations on the committee include the U.S. Army Corps of Engineers, the Environmental Security Technology Certification Program, the Strategic Environmental Research and Development Program, the Institute for Defense Analysis, the U.S. Army Environmental Command, and the U.S. Army Aberdeen Test Center.

**15. SUBJECT TERMS**
Environmental Security Technology Certification Program, APG, Standardized UXO Technology Demonstration Site, blind grid, open field, indirect fire, pushcart.

| **16. SECURITY CLASSIFICATION OF:** | | | **17. LIMITATION OF ABSTRACT** | **18. NUMBER OF PAGES** | **19a. NAME OF RESPONSIBLE PERSON** |
|---|---|---|---|---|---|
| **a. REPORT** | **b. ABSTRACT** | **c. THIS PAGE** | | | **19b. TELEPHONE NUMBER** *(Include area code)* |
| Unclassified | Unclassified | Unclassified | SAR | | |

# ACKNOWLEDGMENTS

Authors:

Leonard Lombardo
Threat Detection and Systems Survivability (TDSS)
U.S. Army Aberdeen Test Center (ATC)
Aberdeen Proving Ground (APG)

Rick Fling
Aberdeen Test Support Services (ATSS)
Sverdrup Technology, Inc.
Aberdeen Proving Ground (APG)

**Contributors:**

Scott Hill
Threat Detection and Systems Survivability (TDSS)
U.S. Army Aberdeen Test Center
Aberdeen Proving Ground

**TABLE OF CONTENTS**

## SECTION 4. TECHNICAL PERFORMANCE RESULTS

**PAGE**



## SECTION 5. APPENDIXES

## SECTION 1. GENERAL INFORMATION

### 1.1 BACKGROUND

Technologies under development for the detection and discrimination of military munitions (MM) (i.e., unexploded ordnance (UXO) and discarded military munitions (DMM)) require testing so that performance can be characterized. To that end, Standardized Test Sites have been developed at Aberdeen Proving Ground (APG), Maryland, and U.S. Army Yuma Proving Ground (YPG), Arizona. These test sites provide a diversity of geology, climate, terrain, and weather as well as diversity in munitions and clutter. Testing at these sites is independently administered and analyzed by the government for the purposes of characterizing technologies, tracking performance with system development, comparing performance of different systems, and comparing performance in different environments (app E, ref 1).

The Standardized UXO Technology Demonstration Site Program is a multiagency program spearheaded by the U.S. Army Environmental Command (USAEC). The U.S. Army Aberdeen Test Center (ATC) and the U.S. Army Corps of Engineers Engineering Research and Development Center (ERDC) provide programmatic support. The program is being funded and supported by the Environmental Security Technology Certification Program (ESTCP), the Strategic Environmental Research and Development Program (SERDP), and the U.S. Army Environmental Quality Technology (EQT) Program.

The Small Munitions Test Site is intended to test the ability of systems to detect anomalies and to discriminate between small ordnance (20-, 37-, and 40-mm projectiles) and clutter when small clutter near the surface partially obstructs the signal. The site is in the form of a grid with a portion of the grid locations having the surface clutter and a portion of the grid locations not having the surface clutter to allow for a comparison between conditions.

### 1.2 SCORING OBJECTIVES

The objective in the Standardized UXO Technology Demonstration Site Program is to evaluate the detection and discrimination capabilities of a given technology under various field and soil conditions. Inert munitions and clutter items are positioned in various orientations and depths in the ground.

The evaluation objectives are as follows:

a. To determine detection and discrimination effectiveness under realistic scenarios with various targets, geology, clutter, density, topography, and vegetation.

b. To determine cost, time, and workforce requirements to operate the technology.

c. To determine the demonstrator’s ability to analyze survey data in a timely manner and provide prioritized Target Lists with associated confidence levels.

d. To provide independent site management to enable the collection of high quality, ground-truth (GT), geo-referenced data for post-demonstration analysis.

### 1.2.1 <u>Scoring Methodology</u>

a. The scoring of the demonstrator's performance is conducted in two stages: response stage and discrimination stage. For both stages, the probability of detection ($P_d$) and the false alarms are reported as receiver-operating characteristic (ROC) curves. False alarms are divided into those anomalies that correspond to emplaced clutter items, measuring the probability of clutter detection ($P_{cd}$) or the probability of false positive ($P_{fp}$). Those that do not correspond to any known item are termed background alarms. The background alarms are addressed as either probability of background alarm ($P_{ba}$) or background alarm rate (BAR).

b. The response stage scoring evaluates the ability of the system to detect emplaced targets without regard to ability to discriminate munitions from other anomaly sources. For the blind grid response stage, the demonstrator provides a target response from each and every grid square along with a threshold below which target responses are deemed insufficient to warrant further investigation. This list is generated with minimal processing and, since a value is provided for every grid square, includes amplitudes both above and below the system noise level. For the open field, the demonstrator provides a list of all anomalies deemed to exceed a demonstrator selected target detection threshold. An item (either munition or clutter) is counted as detected if a demonstrator indicates an anomaly within a specified distance (Halo Radius ($R_{halo}$)) of a ground truth item.

c. The discrimination stage evaluates the demonstrator's ability to correctly identify munitions as such and to reject clutter. For the blind grid discrimination stage, the demonstrator provides the output of the discrimination stage processing for each grid square. For the open field, the demonstrator provides the output of the discrimination stage processing for anomaly reported in the response stage. The values in these lists are prioritized based on the demonstrator's determination that a location is likely to contain munitions. Thus, higher output values are indicative of higher confidence that a munitions item is present at the specified location. For digital signal processing, priority ranking is based on algorithm output. For other discrimination approaches, priority ranking may be based on rule sets or human judgment. The demonstrator also specifies the threshold in the prioritized ranking that provides optimum performance, (i.e., that is expected to retain all detected munitions and reject the maximum amount of clutter).

d. The demonstrator is also scored on efficiency and rejection ratios, which measure the effectiveness of the discrimination stage processing. The goal of discrimination is to retain the greatest number of munitions detections from the anomaly list, while rejecting the maximum number of anomalies arising from nonmunitions items. Efficiency measures the fraction of detected munitions retained after discrimination, while the rejection ratio measures the fraction of false alarms rejected. Both measures are defined relative to the maximum number of munitions detectable by the sensor and its accompanying clutter detection/false positive rate or probability of background alarm.

e. All scoring factors are generated utilizing the Standardized UXO Probability and Plot Program, version 4.

### 1.2.2 Scoring Factors

Factors to be measured and evaluated as part of this demonstration include:

a. Response stage ROC curves:

(1) Probability of detection ($P_d^{res}$).

(2) Probability of clutter detection ($P_{cd}$).

(3) Probability of background alarm ($P_{ba}^{res}$).

b. Discrimination stage ROC curves:

(1) Probability of detection ($P_d^{disc}$).

(2) Probability of false positive ($P_{fp}$).

(3) Probability of background alarm ($P_{ba}^{disc}$).

c. Metrics:

(1) Efficiency (E).

(2) False positive rejection rate ($R_{fp}$).

(3) Background alarm rejection rate ($R_{ba}$).

d. Other:

(1) Probability of detection by size, depth, and density.

(2) Classification by type (i.e., 20-, 40-, 105-mm, etc.).

(3) Location accuracy for single munitions.

(4) Equipment setup, calibration time, and corresponding worker-hour requirements.

(5) Survey time and corresponding worker-hour requirements.

(6) Reacquisition/resurvey time and worker-hour requirements (if any).

(7) Downtime due to system malfunctions and maintenance requirements.

# SECTION 2.   DEMONSTRATION

## 2.1   DEMONSTRATOR INFORMATION

### 2.1.1   Demonstrator Point of Contact (POC) and Address

POC:        Jeff Gamey

Address:    100A Donner Drive, Oak Ridge, TN   37830

### 2.1.2   System Description (Provided by Demonstrator)

The TEMTADS (TEM)-8g is a time-domain electromagnetic array consisting of a single Z-axis transmitter and eight Z-axis receivers, towed on a wheeled or skid-mounted platform behind a utility vehicle.  Navigation is visual for small areas (all APG sites).  Data positioning uses a dual Differential Global Positioning System (DGPS) with post-processing to determine location and orientation.

The transmitter uses an alternating castle waveform with a 30-Hz base frequency and a 50-percent duty cycle.  The transmitter frame is 2 by 0.75 m with 12 turns of wire carrying 60 A of current and producing a peak magnetic moment of 1080 $Am^2$.  The receivers are 0.2-m diameter circles spaced at 0.22-m intervals in a line across the middle of the transmitter.  Sensor height is variable and positioned as low as possible to the ground.  Nominal sensor height is 15 to 25 cm.

Data are recorded at 30 Hz for seven geometrically spaced time-gates between 0.4 and 8 ms.  At a forward speed of approximately 2 m/s, this represents a down-line data spacing of 0.07 m.  A single-pass swath covers 1.8 m.  A line spacing of 1.5 m (or 5 ft.) is used to ensure complete coverage of the site.

Figure 1.   TEM-8G/TOWED ARRAY .

### 2.1.3 **Data Processing Description (Provided by Demonstrator)**

The section should be submitted for each area surveyed by the vendor. Discussion should include how target selection, parameter estimation, and classification vary by site area and objective. The following information should be submitted to ATC within 30 days before each area is surveyed:

a. Target Selection Criteria. This section will detail the target selection criteria and the data required to implement the criteria by answering the following questions:

(1) What kind of preprocessing (if any) is applied to the raw data (e.g., filtering, etc.)?

Minimal low-pass filter.

(2) What is the format of the data both pre- and post-processing of the raw data (e.g., ASCII, binary, etc.)?

Recorded as binary, converted to ASCII and imported to Geosoft GDB.

(3) What algorithm is used for detection (e.g., peaks of signal surpassing threshold, etc.)?

Peakedness, with proximate anomalies combined to a single target.

(4) Why is this algorithm used and not others?

Past experience has shown this to be a reliable method.

(5) On what principles is the algorithm based (e.g., statistical models, heuristic rules, etc.)?

Tests each grid point relative to those immediately adjacent to it, and outward in increasing circles in order to find reliable peak locations.

(6) What tunable parameters (if any) are used in the detection process (e.g., threshold on signal amplitude, window length, filter coefficients, etc.)?

The number of circles of data tested and the number of points in each ring which the center value must be larger than can be adjusted.

(7) What are the final values of all tunable parameters for the detection algorithm?

This will be determined by analysis of the Calibration Grid data.

b. Parameter Estimation. This section should include the details of which parameters will be extracted from the sensor data for each detected item for characterization. Please answer the following questions:

(1) Which characteristics will be extracted from each detected item and input to the discrimination algorithm (e.g., depth, size, polarizability coefficients, fit quality, etc.)?

Output parameters include target location, depth, inversion fit quality and polarizability decays.

(2) Why have these characteristics been chosen and not others (e.g., empirical evidence of their ability to help discriminate, inclusion in a theoretical tradition, etc.)?

These parameters are the industry standard for ordnance classification.

(3) How are these characteristics estimated (e.g., least-mean-squares fit to a dipole model, etc.), include the equations that are used for parameter estimation?

They are estimated from a least squares fit to a dipole model.

(4) What tunable parameters (if any) are used in the characterization process (e.g., thresholds on background noise, etc.)?

The number of dipoles and the background zero levels can be adjusted to improve inversion results.

c. Classification. This section should include the details describing the algorithm and associated data and parameters used for discrimination by answering the following questions:

(1) What algorithm is used for discrimination (e.g., multi-layer perception, support vector machine, etc.)?

A rules-based classification system has been derived from training data sets.

(2) Why is this algorithm used and not others?

This system has proven reliable on other calibration targets.

(3) Which parameters are considered as possible inputs to the algorithm?

The amplitude of the primary polarizability, the amplitude of the secondary polarizability, and the decay of the primary polarizability are the three input parameters. These represent a three-dimensional parameter space for classification. Measured results are compared to a library of target types. If the measured results are close enough to the average library values then an ordnance declaration is made.

(4) What are the outputs of the algorithm (probabilities, confidence levels)?

The output is a dig list divided into four basic categories: Targets-of interest (TOI), non-TOI, cannot decide, cannot analyze.

(5) How is the threshold set to decide where the munitions/nonmunitions line lies in the discrimination process?

Thresholds are set based on the standard deviation of the training sets.

d. Training. This section should include the details of how training data are used to make a decision on the likelihood of the anomaly correspondence to munitions. Please answer the following questions:

(1) Which tunable parameters have final values that are optimized over a training set of data and which have values that are set according to geophysical knowledge (i.e., intuition, experience, common sense)?

The average library values for each target are based on a measured training set. The thresholds, in terms of the number of standard deviations within which to make a declaration, are based on a combination of the training data and experience with the measured values from a particular site.

(a) For those tunable parameters with final values set according to geophysical knowledge:

1 What is the reasoning behind choosing these particular values?

2 Why were the final values not optimized over a training set of data?

This is a new system and the size of the training set is still too small.

(b) For those tunable parameters with final values optimized over the training set data:

1 What training data is used (e.g., all data, a randomly chosen portion of data, etc.)?

All targets from the Calibration Grid are used.

2 What error metric is minimized during training (e.g., mean squared error, etc.)?

Results are averaged and standard deviation calculated.

3 What learning rule is used during training (e.g., gradient descent, etc.)?

Not applicable.

4 What criterion is used to stop training (e.g., number of iterations exceeds threshold, good generalization over validation set of data, etc.)?

Not applicable.

5 Are all tunable parameters optimized at once or in sequence (in sequence = parameters 1 is held constant at some common sense values while parameter 2 is optimized, and then parameter 2 is held constant at its optimized value while parameter 1 is optimized)?

Not applicable.

(2) What are the final values of all tunable parameters for the characterization process?

### 2.1.4 Data Submission Format

Data were submitted for scoring in accordance with data submission protocols outlined on the USAEC Web site www.uxotestsites.org. These submitted data are not included in this report in order to protect GT information.

### 2.1.5 Demonstrator Quality Assurance (QA) and Quality Control (QC) (Provided by Demonstrator)

### 2.1.6 Additional Records

The following record(s) by this vendor can be accessed via the Internet as Microsoft Word documents at www.uxotestsites.org.

## 2.2 APG SITE INFORMATION

### 2.2.1 Location

The APG Standardized Test Site is located within a secured range area of the Aberdeen Area. The Aberdeen Area of APG is located approximately 30 miles northeast of Baltimore at the northern end of the Chesapeake Bay. The Standardized Test Site encompasses 17 acres of upland and lowland flats, woods, and wetlands.

### 2.2.2 Soil Type

According to the soils survey conducted for the entire area of APG in 1998, the test site consists primarily of Elkton Series type soil (ref 2). The Elkton Series consist of very deep, slowly permeable, poorly drained soils. These soils formed in silty aeolin sediments and the underlying loamy alluvial and marine sediments. They are on upland and lowland flats and in depressions of the Mid-Atlantic Coastal Plain. Slopes range from 0 to 2 percent.

ERDC conducted a site-specific analysis in May 2002 (ref 3). The results basically matched the soil survey mentioned above. Seventy percent of the samples taken were classified as silty loam. The majority (77 percent) of the soil samples had a measured water content between 15 and 30 percent with the water content decreasing slightly with depth.

For more details concerning the soil properties at the APG test site, go to www.uxotestsites.org on the Web to view the entire soils description report.

#### 2.2.3 Test Areas

A description of the test site areas at APG associated with the Small Munitions Test Site is presented in Table 1.

**TABLE 1. TEST SITE AREAS**

| Area | Description |
|---|---|
| Small munitions calibration lanes | Contains three standard munitions items buried in six positions at various angles and depths to allow demonstrators to calibrate their equipment. |
| Small munitions grid | Contains 300 grid cells. The center of each grid cell contains either munitions, clutter, or nothing with a portion containing surface clutter. |

### 2.3 ATC SURVEY COMMENTS

None.

## SECTION 3. FIELD DATA

### 3.1 DATE OF FIELD ACTIVITIES (14 THROUGH 16 May 2013)

### 3.2 AREAS TESTED/NUMBER OF HOURS

Areas tested and total numbers of hours operated at each site are presented in Table 2.

**TABLE 2. AREAS TESTED AND NUMBER OF HOURS**

| Area | No. of Hours |
|---|---|
| Calibration lanes | 1.83 |
| Small munitions grid | 9.08 |

Note: Table 2 represents the total time spent in each area.

### 3.3 TEST CONDITIONS

#### 3.3.1 Weather Conditions

An APG weather station located approximately 1 mile west of the test site was used to record average temperature and precipitation on a half-hour basis for each day of operation. The temperatures presented in Table 3 represent the average temperature during field operations from 0700 to 1700 hours, while precipitation data represent a daily total amount of rainfall. Hourly weather logs used to generate this summary are provided in Appendix B.

**TABLE 3. TEMPERATURE/PRECIPITATION DATA SUMMARY**

| Date, 2013 | Average Temperature, ºF | Total Daily Precipitation, in. |
|---|---|---|
| 14 May | 55.1 | 0.02 |
| 15 May | 64.7 | 0.00 |
| 16 May | 74.3 | 0.00 |

#### 3.3.2 Field Conditions

The weather was warm and the field dry throughout the survey period for Battelle.

### 3.3.3 Soil Moisture

Three soil probes were placed at various locations within the site to capture soil moisture data: blind grid, calibration, open field, and wooded areas. Measurements were collected in percent moisture and were taken twice daily (morning and afternoon) from five different soil depths (1 to 6 in., 6 to 12 in., 12 to 24 in., 24 to 36 in., and 36 to 48 in.) from each probe. Soil moisture logs are provided in Appendix C.

## 3.4 FIELD ACTIVITIES

### 3.4.1 Setup/Mobilization

These activities included initial mobilization and daily equipment preparation and breakdown. A three-person crew took 5 hours and 30 minutes to perform the initial setup and mobilization. A total of 2 hours and 10 minutes of equipment preparation was accrued, and end of day equipment breakdown totaled 45 minutes.

### 3.4.2 Calibration

Battelle spent a total of 1 hour and 50 minutes in the calibration lanes, of which 45 minutes were spent collecting data. Several calibration exercises occurred while surveying the Blind Grid and Calibration Grid, totaling 1 hour and 15 minutes.

### 3.4.3 Downtime Occasions

Occasions of downtime are grouped into five categories: equipment/data checks or equipment maintenance, equipment failure and repair, weather, demonstration site issues, or breaks/lunch. All downtime is included for the purposes of calculating labor requirements (section 5) except for downtime due to demonstration site issues. Demonstration site issues, while noted in the daily log, are considered nonchargeable downtime for the purposes of calculating labor costs and are not discussed. Breaks and lunches are discussed in this section and billed to the total site survey area.

**3.4.3.1 Equipment/data checks, maintenance.** Equipment data checks and maintenance activities accounted for no site usage time. These activities included changing out batteries and performing routine data checks to ensure the data were being properly recorded/collected. Battelle spent 35 minutes for breaks and lunches.

**3.4.3.2 Equipment failure or repair.** No equipment failure or repair occurred during this survey.

**3.4.3.3 Weather.** One weather delay occurred during the survey. A brief shower on 14 May 2013 curtailed survey activities for 10 minutes.

#### 3.4.4 <u>Data Collection</u>

**TABLE 5. TOTAL TIME BATTELLE, SPENT PER AREA**

| Area | Time |
|---|---|
| Blind grid | - |
| Open field | - |
| Legacy | - |
| Direct fire | - |
| Indirect fire | - |
| Challenge | - |
| Wooded | - |
| Mine Grid | - |
| Moguls | - |
| Small munition | 4 hr, 40 min |

Note: Table 5 represents the total time spent in each area collecting data.

#### 3.4.5 <u>Demobilization</u>

The Battelle survey crew conducted a demonstration of the calibration and blind grid. Demobilization occurred on 16 May 2013. On that day, it took the crew 2 hours and 5 minutes to break down and pack up their equipment.

### 3.5 PROCESSING TIME

Battelle submitted the raw data from the demonstration activities on the last day of the demonstration, as required. The scoring submittal data were provided in August 2013.

### 3.6 DEMONSTRATOR'S FIELD PERSONNEL

Jeff Gamey
Jeanmarie Norton
Kent Vaught

### 3.7 DEMONSTRATOR'S FIELD SURVEYING METHOD

Battelle surveyed the area in a linear fashion. They utilized line spacing of 1.5 meters.

## 3.8 SUMMARY OF DAILY LOGS

Daily logs capture all field activities during this demonstration and are provided in Appendix D.

## SECTION 4. TECHNICAL PERFORMANCE RESULTS

### 4.1 ROC CURVES USING ALL MUNITIONS CATEGORIES

The probability of detection for the response stage ($P_d^{res}$) and the discrimination stage ($P_d^{disc}$) versus their respective probability of clutter detection or probability of false positive for capped and uncapped test areas are shown in Figures 2 and 3, respectively. The probabilities plotted against their respective BAR for capped and uncapped conditions are shown in Figures 4 and 5, respectively. All figures use horizontal lines to illustrate the performance of the demonstrator at two demonstrator-specified points: at the system noise level for the response stage, representing the point below which targets are not considered detectable, and at the demonstrator's recommended threshold level for the discrimination stage, defining the subset of targets the demonstrator would recommend digging based on discrimination.

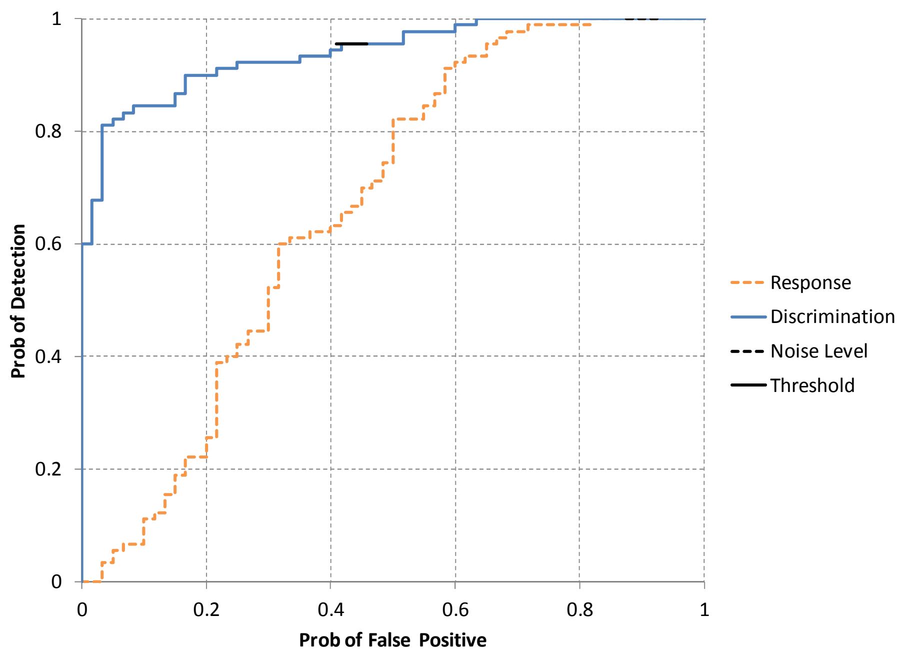


Figure 2. TEM-8G/TOWED ARRAY capped test area probability of detection for response and discrimination stages versus their respective probability of false positive.

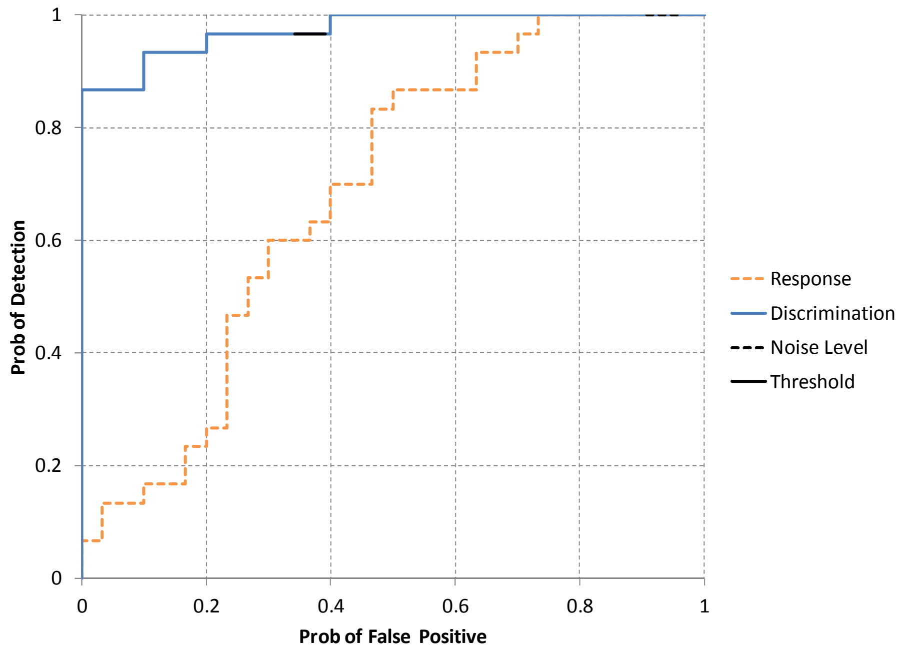


Figure 3. TEM-8G/TOWED ARRAY uncapped test area probability of detection for response and discrimination stages versus their respective probability of false positive.

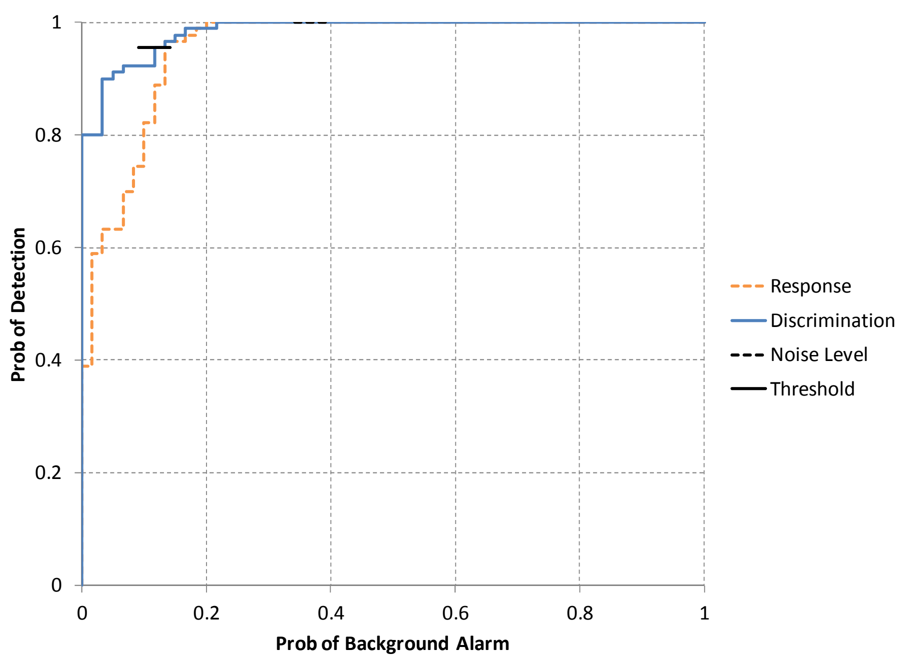


Figure 4. TEM-8G/TOWED ARRAY capped test area probability of detection for response and discrimination stages versus their respective probability of background alarm.

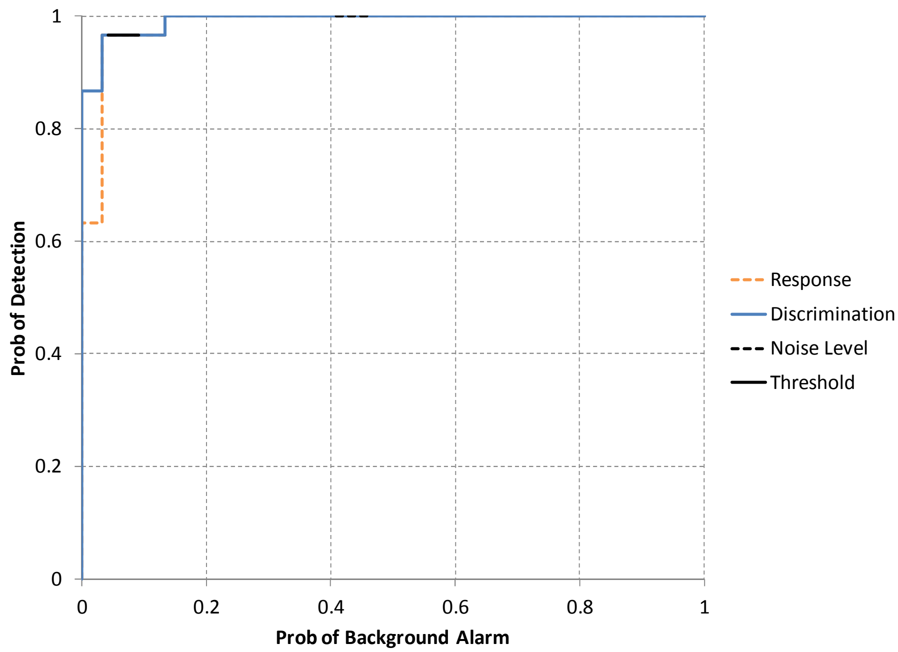


Figure 5. TEM-8G/TOWED ARRAY uncapped test area probability of detection for response and discrimination stages versus their respective probability of background alarm.

## 4.2 ROC CURVES USING ORDNANCE LARGER THAN 20 MM

The $P_d$ for the response stage ($P_d^{res}$) and the discrimination stage ($P_d^{disc}$) versus their respective $P_{cd}$ or $P_{fp}$ for capped and uncapped test areas when only ordnance larger than 20 mm are scored are shown in Figures 6 and 7, respectively. The probabilities plotted against their respective BAR for capped and uncapped conditions when only ordnance larger than 20 mm are scored are shown in Figures 8 and 9, respectively. All figures use horizontal lines to illustrate the performance of the demonstrator at two demonstrator-specified points: at the system noise level for the response stage, representing the point below which targets are not considered detectable, and at the demonstrator's recommended threshold level for the discrimination stage, defining the subset of targets the demonstrator would recommend digging based on discrimination.

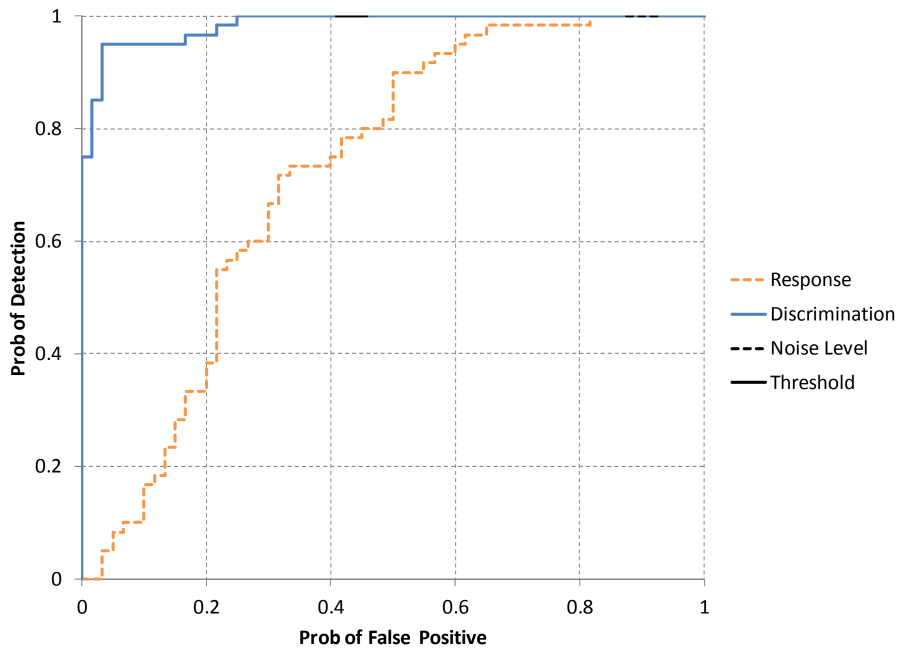


Figure 6. TEM-8G/TOWED ARRAY capped test area probability of detection for response and discrimination stages versus their respective probability of false positive for all ordnance larger than 20 mm.

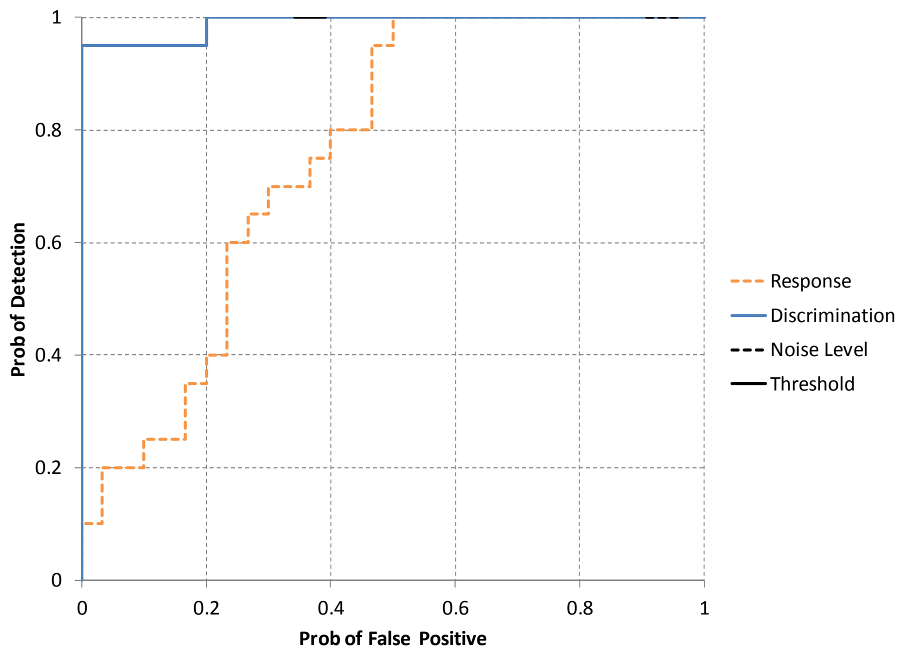


Figure 7. TEM-8G/TOWED ARRAY uncapped test area probability of detection for response and discrimination stages versus their respective probability of false positive for all ordnance larger than 20 mm.

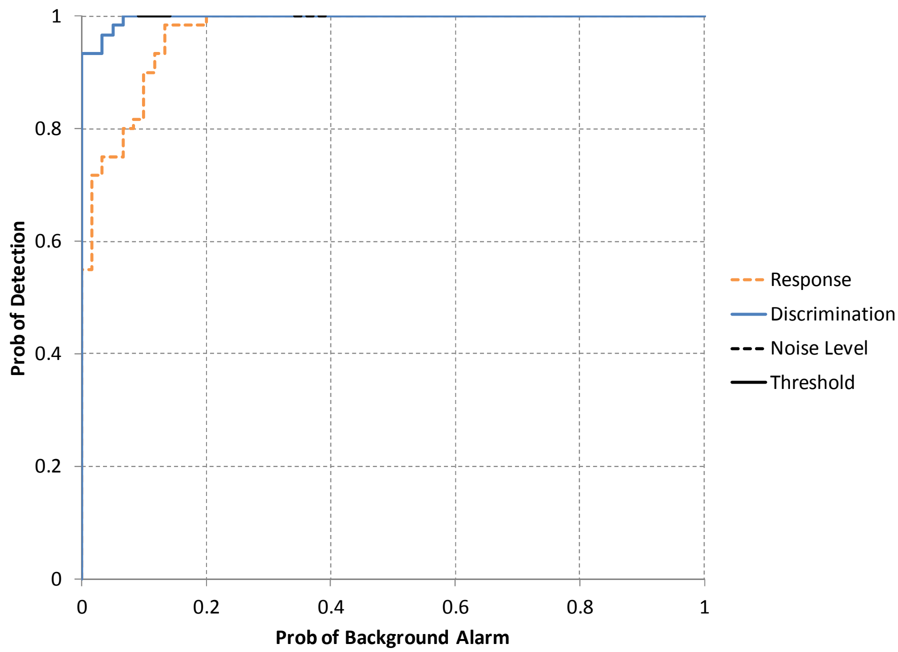


Figure 8. TEM-8G/TOWED ARRAY capped test area probability of detection for response and discrimination stages versus their respective probability of background alarm for all ordnance larger than 20 mm.

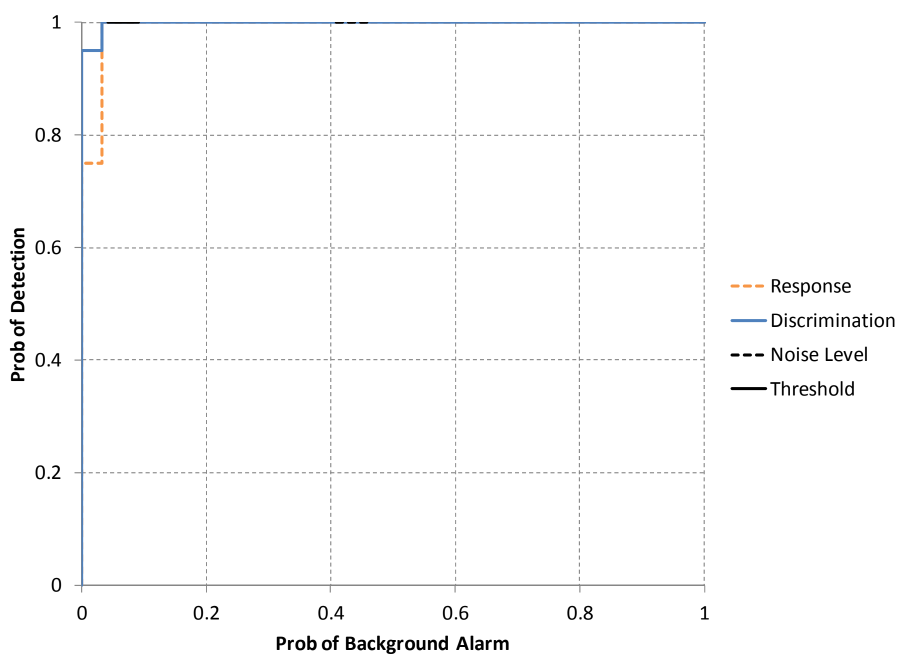


Figure 9. TEM-8G/TOWED ARRAY uncapped test area probability of detection for response and discrimination stages versus their respective probability of background alarm for all ordnance larger than 20 mm.

## 4.3 PERFORMANCE SUMMARIES

Results for capped and uncapped test areas are presented in Tables 4a and 4b, respectively. The response stage results were derived from the list of anomalies above the demonstrator-provided noise level. The results for the discrimination stage were derived from the demonstrator's recommended threshold for optimizing munitions related cleanup by minimizing false alarm digs and maximizing munitions recovery. The lower and upper 90-percent confidence limits on $P_d$, $P_{cd}$, and $P_{fp}$ were calculated assuming that the number of detections and false positives are binomially distributed random variables.

### TABLE 4a. CAPPED TEST AREA RESULTS

| | Response Stage | | | | Discrimination Stage | | | |
|---|---|---|---|---|---|---|---|---|
| Munitions[a] Scores | ***$P_d^{res}$: by type*** | | | | ***$P_d^{disc}$: by type*** | | | |
| | **All Types** | **40-mm** | **37-mm** | **20-mm** | **All Types** | **40-mm** | **37-mm** | **20-mm** |
| | 1.000 | 1.000 | 1.000 | 1.000 | 0.980 | 1.000 | 1.000 | 0.941 |
| | 1.000 | 1.000 | 1.000 | 1.000 | 0.956 | 1.000 | 1.000 | 0.867 |
| | 0.975 | 0.926 | 0.926 | 0.926 | 0.913 | 0.926 | 0.926 | 0.751 |
| *By Depth[b]* | | | | | | | | |
| **0 to 4D** | 1.000 | 1.000 | NA | NA | 1.000 | 1.000 | NA | NA |
| **4D to 8D** | 1.000 | 1.000 | 1.000 | 1.000 | 0.943 | 1.000 | 1.000 | 0.778 |
| **8D to 12D** | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| **> 12D** | 1.000 | 1.000 | 1.000 | 1.000 | 0.867 | 1.000 | 1.000 | 0.714 |

| Clutter Scores | ***$P_{cd}$*** | | | ***$P_{fp}$*** | | |
|---|---|---|---|---|---|---|
| *By Mass* | | | | | | |
| *By Depth[b]* | **All Mass** | **0 to 0.25 kg** | **>0.25 to 1 kg** | **All Mass** | **0 to 0.25 kg** | **>0.25 to 1 kg** |
| **All Depth** | 0.947 | 0.902 | 1.000 | 0.524 | 0.580 | 0.549 |
| | 0.900 | 0.818 | 1.000 | 0.433 | 0.455 | 0.407 |
| | 0.831 | 0.703 | 0.918 | 0.346 | 0.333 | 0.277 |
| **0 to 0.15 m** | 1.000 | NA | 1.000 | 0.333 | NA | 0.333 |
| **0.15 to 0.3 m** | 1.000 | 1.000 | 1.000 | 0.548 | 0.667 | 0.474 |
| **0.3 to 0.6 m** | 0.739 | 0.714 | 1.000 | 0.304 | 0.333 | 0.000 |
| Background Alarm Rates | | | | | | |
| | **$P_{ba}^{res}$:** 0.367 | | | **$P_{ba}^{disc}$:** 0.117 | | |

[a]The two numbers to the right of the all types munitions result are an upper and lower 90-percent confidence interval for an assumed binomial distribution.
[b]All depths are measured to the center of the object.

## TABLE 4b. UNCAPPED TEST AREA RESULTS

| | Response Stage | | | | Discrimination Stage | | | |
|---|---|---|---|---|---|---|---|---|
| Munitions[a] Scores | ***$P_d^{res}$: by type*** | | | | ***$P_d^{disc}$: by type*** | | | |
| | **All Types** | **40-mm** | **37-mm** | **20-mm** | **All Types** | **40-mm** | **37-mm** | **20-mm** |
| | 1.000<br>1.000<br>0.926 | 1.000<br>1.000<br>0.794 | 1.000<br>1.000<br>0.794 | 1.000<br>1.000<br>0.794 | 0.996<br>0.967<br>0.876 | 1.000<br>1.000<br>0.794 | 1.000<br>1.000<br>0.794 | 0.990<br>0.900<br>0.663 |
| *By Depth[b]* | | | | | | | | |
| **0 to 4D** | 1.000 | NA | 1.000 | 1.000 | 1.000 | NA | 1.000 | 1.000 |
| **4D to 8D** | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| **8D to 12D** | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| **> 12D** | 1.000 | 1.000 | 1.000 | 1.000 | 0.800 | 1.000 | 1.000 | 0.500 |

| Clutter Scores | ***$P_{cd}$*** | | | ***$P_{fp}$*** | | |
|---|---|---|---|---|---|---|
| | *By Mass* | | | | | |
| *By Depth[b]* | **All Mass** | **0 to 0.25 kg** | **>0.25 to 1 kg** | **All Mass** | **0 to 0.25 kg** | **>0.25 to 1 kg** |
| **All Depth** | 0.982<br>0.933<br>0.832 | 0.994<br>0.941<br>0.790 | 0.992<br>0.923<br>0.732 | 0.500<br>0.367<br>0.248 | 0.478<br>0.294<br>0.151 | 0.669<br>0.462<br>0.264 |
| **0 to 0.15 m** | 1.000 | NA | 1.000 | 0.500 | NA | 0.500 |
| **0.15 to 0.3 m** | 1.000 | 1.000 | 1.000 | 0.563 | 0.571 | 0.556 |
| **0.3 to 0.6 m** | 0.833 | 0.900 | 0.500 | 0.083 | 0.100 | 0.000 |
| Background Alarm Rates | | | | | | |
| | **$P_{ba}^{res}$:** 0.433 | | | **$P_{ba}^{disc}$:** 0.067 | | |

[a]The two numbers to the right of the all types munitions result are an upper and lower 90-percent confidence interval for an assumed binomial distribution.
[b]All depths are measured to the center of the object.

## 4.4 EFFICIENCY, REJECTION RATES, AND TYPE CLASSIFICATION

Efficiency and rejection rates are calculated to quantify the discrimination ability at specific points of interest on the ROC curve: (1) at the point where no decrease in $P_d$ is suffered (i.e., the efficiency is by definition equal to one) and (2) at the operator selected threshold. These values are presented in Tables 5a and 5b.

## TABLE 5a. CAPPED EFFICIENCY AND REJECTION RATES

| | **Efficiency (E)** | **False Positive Rejection Rate** | **Background Alarm Rejection Rate** |
|---|---|---|---|
| At Operating Point | 0.96 | 0.52 | 0.68 |
| With No Loss of $P_d$ | 1.00 | 0.30 | 0.41 |

TABLE 5b. UNCAPPED EFFICIENCY AND REJECTION RATES

| | Efficiency (E) | False Positive Rejection Rate | Background Alarm Rejection Rate |
|---|---|---|---|
| At Operating Point | 0.97 | 0.61 | 0.85 |
| With No Loss of $P_d$ | 1.00 | 0.57 | 0.69 |

At the demonstrator's recommended setting, the munitions items that were detected and correctly discriminated were further scored on whether their correct type could be identified (tables 6a and 6b). Correct type examples include 20-mm projectile, 37-mm projectile, and 40-mm projectile. A list of the standard type declaration required for each munitions item was provided to demonstrators prior to testing. The standard types for the three example items are 20-mmP, 37-mmP, and 40-mmP.

TABLE 6a. CAPPED CORRECT TYPE CLASSIFICATION OF TARGETS CORRECTLY DISCRIMINATED AS MUNITIONS

| Size | Percentage Correct |
|---|---|
| 20-mm | 0.77 |
| 37-mm | 0.97 |
| 40-mm | 1.00 |
| Overall | 0.91 |

TABLE 6b. UNCAPPED CORRECT TYPE CLASSIFICATION OF TARGETS CORRECTLY DISCRIMINATED AS MUNITIONS

| Size | Percentage Correct |
|---|---|
| 20-mm | 0.90 |
| 37-mm | 1.00 |
| 40-mm | 1.00 |
| Overall | 0.97 |

## 4.5 LOCATION ACCURACY

The mean location error and standard deviations are presented in Tables 7a and 7b. These calculations are based on average missed distance for munitions correctly identified during the response stage. Depths are measured from the center of the munitions to the surface. For the blind grid, only depth errors are calculated because (X, Y) positions are known to be the centers of the grid square.

**TABLE 7a. CAPPED MEAN LOCATION ERROR AND STANDARD DEVIATION**

| | Mean | Standard Deviation |
|---|---|---|
| Depth | 0.050 | 0.040 |

**TABLE 7b. UNCAPPED MEAN LOCATION ERROR AND STANDARD DEVIATION**

| | Mean | Standard Deviation |
|---|---|---|
| Depth | 0.041 | 0.027 |

## 4.6 STATISTICAL COMPARISONS

Statistical two-sided Fisher's exact significance tests were used to compare results between the capped and uncapped test areas. The intent of the comparison is to determine if the cap has an effect on the performance of the sensor system. However, any modifications in the UXO sensor system during the test, like changes in the processing or changes in the selection of the operating threshold, will also contribute to performance differences.

The Fisher's exact test was used to compare capped to uncapped test areas with regard to $P_d^{res}$, $P_d^{disc}$, $P_{fp}^{res}$, $P_{fp}^{disc}$, $P_{ba}^{res}$, $P_{ba}^{disc}$, Efficiency, False Positive Rejection Rate ($R_{fp}$) and Background Alarm Rejection Rate ($R_{ba}$). The confidence level used for the test was 0.1. The test results are presented in Table 8. A detailed explanation and example of the Fisher's exact test application is located in Appendix A.

## TABLE 8. FISHER'S EXACT RESULTS - CAPPED VERSUS UNCAPPED

| Metric | Ordnance Size | | | Overall |
|---|---|---|---|---|
| | 20-mm | 37-mm | 40-mm | |
| $P_d^{res}$ | Not significant | Not significant | Not significant | Not significant |
| $P_d^{disc}$ | Not significant | Not significant | Not significant | Not significant |
| $P_{fp}^{res}$ | | | | Not significant |
| $P_{fp}^{disc}$ | | | | Not significant |
| $P_{ba}^{res}$ | | | | Not significant |
| $P_{ba}^{disc}$ | | | | Not significant |
| Efficiency | | | | Not significant |
| $R_{FP}$ | | | | Not significant |
| $R_{BA}$ | | | | Not significant |

# SECTION 5. APPENDIXES

## APPENDIX A. TERMS AND DEFINITIONS

GENERAL DEFINITIONS

Anomaly: Location of a system response deemed to warrant further investigation by the demonstrator for consideration as an emplaced munitions item.

Detection: An anomaly location that is within $R_{halo}$ of an emplaced munitions item.

Military Munitions (MM): Specific categories of MM that may pose unique explosive safety risks, including UXO as defined in 10 USC 101(e)(5), DMM as defined in 10 USC 2710(e)(2) and/or munitions constituents (e.g., TNT, RDX) as defined in 10 USC 2710(e)(3) that are present in high enough concentrations to pose an explosive hazard.

Emplaced Munitions: A munitions item buried by the government at a specified location in the test site.

Emplaced Clutter: A clutter item (i.e., nonmunitions item) buried by the government at a specified location in the test site.

$R_{halo}$: A predetermined radius about an emplaced item (clutter or munitions) within which an anomaly identified by the demonstrator as being of interest is considered to be a detection of that item. For the purpose of this program, a circular halo 0.5 meter in radius is placed around the center of the object for all clutter and munitions items.

Small Munitions: Caliber of munitions less than or equal to 40 mm (includes 20-mm projectile, 25-mm projectile, 37-mm projectile, 40-mm projectile, submunitions BLU-26, BLU-63, and M42).

Medium Munitions: Caliber of munitions greater than 40 mm and less than or equal to 81 mm (includes 57-mm projectile, 60-mm mortar, 2.75-inch rocket, and 81-mm mortar).

Large Munitions: Caliber of munitions greater than 81 mm (includes 105-mm high-explosive antitank (HEAT), 105-mm projectile, and 155-mm projectile).

Group: Two or more adjacent GT items with overlapping halos.

GT: Ground truth

Response Stage Noise Level: The level that represents the signal level below which anomalies are not considered detectable. Demonstrators are required to provide the recommended noise level for the blind grid test area.

Discrimination Stage Threshold: The demonstrator-selected threshold level that is expected to provide optimum performance of the system by retaining all detectable munitions and rejecting the maximum amount of clutter. This level defines the subset of anomalies the demonstrator would recommend digging based on discrimination.

Binomially Distributed Random Variable: A random variable of the type which has only two possible outcomes, say success and failure, is repeated for $n$ independent trials with the probability $p$ of success and the probability $1-p$ of failure being the same for each trial. The number of successes $x$ observed in the $n$ trials is an estimate of $p$ and is considered to be a binomially distributed random variable.

RESPONSE AND DISCRIMINATION STAGE DATA

The scoring of the demonstrator's performance is conducted in two stages: response stage and discrimination stage. For both stages, the probability of detection ($P_d$) and the false alarms are reported as receiver-operating characteristic (ROC) curves. False alarms are divided into those anomalies that correspond to emplaced clutter items, measuring the probability of clutter detection ($P_{cd}$) or probability of false positive ($P_{fp}$). Those that do not correspond to any known item are termed background alarms.

The response stage is a measure of whether the sensor can detect an object of interest. For a channel instrument, this value should be closely related to the amplitude of the signal. The demonstrator must report the response level (threshold) below which target responses are deemed insufficient to warrant further investigation. At this stage, minimal processing may be done. This includes filtering long- and short-scale variations, bias removal, and scaling. This processing should be detailed in the data submission.

For a multichannel instrument, the demonstrator must construct a quantity analogous to amplitude. The demonstrator should consider what combination of channels provides the best test for detecting any object that the sensor can detect. The average amplitude across a set of channels is an example of an acceptable response stage quantity. Other methods may be more appropriate for a given sensor. Again, minimal processing can be done, and the demonstrator should explain how this quantity was constructed in their data submission.

The discrimination stage evaluates the demonstrator's ability to correctly identify munitions as such, and to reject clutter. For the same locations as in the response stage anomaly list, the discrimination stage list contains the output of the algorithms applied in the discrimination-stage processing. This list is prioritized based on the demonstrator's determination that an anomaly location is likely to contain munitions. Thus, higher output values are indicative of higher confidence that a munitions item is present at the specified location. For electronic signal processing, priority ranking is based on algorithm output. For other systems, priority ranking is based on human judgment. The demonstrator also selects the threshold that the demonstrator believes will provide optimum system performance, (i.e., that retains all the detected munitions and rejects the maximum amount of clutter).

Note: The two lists provided by the demonstrator contain identical numbers of potential target locations. They differ only in the priority ranking of the declarations.

RESPONSE STAGE DEFINITIONS

Response Stage Probability of Detection ($P_d^{res}$): $P_d^{res}$ = (No. of response-stage detections)/(No. of emplaced munitions in the test site).

Response Stage Clutter Detection ($cd^{res}$): An anomaly location that is within $R_{halo}$ of an emplaced clutter item.

Response Stage Probability of Clutter Detection ($P_{cd}^{res}$): $P_{cd}^{res}$ = (No. of response-stage clutter detections)/(No. of emplaced clutter items).

Response Stage Background Alarm ($ba^{res}$): An anomaly in a blind grid cell that contains neither emplaced munitions nor an emplaced clutter item. An anomaly location in the open field or scenarios that is outside $R_{halo}$ of any emplaced munitions or emplaced clutter item.

Response Stage Probability of Background Alarm ($P_{ba}^{res}$): Blind grid only: $P_{ba}^{res}$ = (No. of response-stage background alarms)/(No. of empty grid locations).

Response Stage Background Alarm Rate ($BAR^{res}$): Open field any challenge area (including the direct and indirect firing sub areas) only: $BAR^{res}$ = (No. of response-stage background alarms)/(arbitrary constant).

Note that the quantities $P_d^{res}$, $P_{cd}^{res}$, $P_{ba}^{res}$, and $BAR^{res}$ are functions of $t^{res}$, the threshold applied to the response-stage signal strength. These quantities can therefore be written as $P_d^{res}(t^{res})$, $P_{cd}^{res}(t^{res})$, $P_{ba}^{res}(t^{res})$, and $BAR^{res}(t^{res})$.

DISCRIMINATION STAGE DEFINITIONS

Discrimination: The application of a signal processing algorithm or human judgment to sensor data to discriminate munitions from clutter. Discrimination should identify anomalies that the demonstrator has high confidence correspond to munitions, as well as those that the demonstrator has high confidence correspond to nonmunitions or background returns. The former should be ranked with highest priority and the latter with lowest.

Discrimination Stage Probability of Detection ($P_d^{disc}$): $P_d^{disc}$ = (No. of discrimination-stage detections)/(No. of emplaced munitions in the test site).

Discrimination Stage False Positive ($fp^{disc}$): An anomaly location that is within $R_{halo}$ of an emplaced clutter item.

Discrimination Stage Probability of False Positive ($P_{fp}^{disc}$): $P_{fp}^{disc}$ = (No. of discrimination stage false positives)/(No. of emplaced clutter items).

Discrimination Stage Background Alarm ($ba^{disc}$): An anomaly in a blind grid cell that contains neither emplaced munitions nor an emplaced clutter item. An anomaly location in the open field or scenarios that is outside $R_{halo}$ of any emplaced munitions or emplaced clutter item.

Discrimination Stage Probability of Background Alarm ($P_{ba}^{disc}$): $P_{ba}^{disc}$ = (No. of discrimination-stage background alarms)/(No. of empty grid locations).

Discrimination Stage Background Alarm Rate ($BAR^{disc}$): $BAR^{disc}$ = (No. of discrimination-stage background alarms)/(arbitrary constant).

Note that the quantities $P_d^{disc}$, $P_{fp}^{disc}$, $P_{ba}^{disc}$, and $BAR^{disc}$ are functions of $t^{disc}$, the threshold applied to the discrimination-stage signal strength. These quantities can therefore be written as $P_d^{disc}(t^{disc})$, $P_{fp}^{disc}(t^{disc})$, $P_{ba}^{disc}(t^{disc})$, and $BAR^{disc}(t^{disc})$.

RECEIVER-OPERATING CHARACTERISTIC (ROC) CURVES

ROC curves at both the response and discrimination stages can be constructed based on the above definitions. The ROC curves plot the relationship between $P_d$ versus $P_{cd}$ or $P_{fp}$ and $P_d$ versus BAR or $P_{ba}$ as the threshold applied to the signal strength is varied from its minimum ($t_{min}$) to its maximum ($t_{max}$) value.[1] $P_d$ versus $P_{fp}$ and $P_d$ versus BAR being combined into ROC curves are shown in Figure A-1. Note that the “res” and “disc” superscripts have been suppressed from all the variables for clarity.

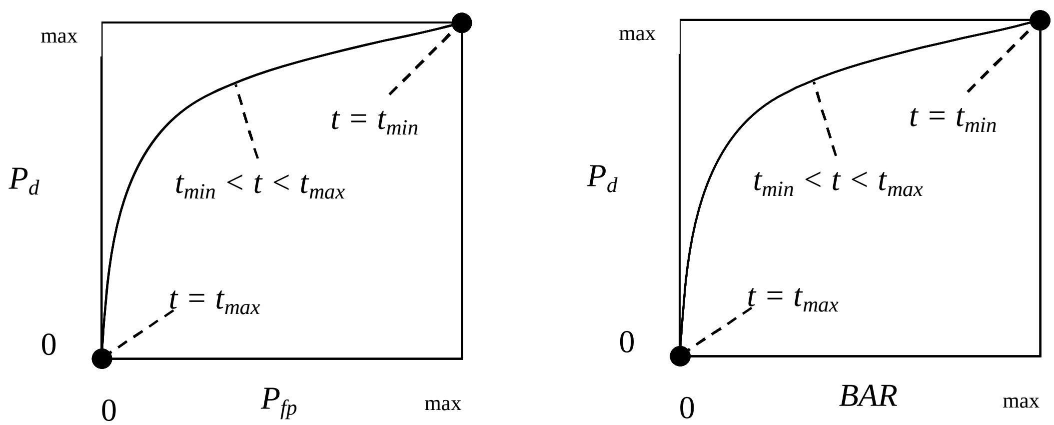


Figure A-1. ROC curves for open field testing. Each curve applies to both the response and discrimination stages.

METRICS TO CHARACTERIZE THE DISCRIMINATION STAGE

The demonstrator is also scored on efficiency and rejection ratio, which measure the effectiveness of the discrimination stage processing. The goal of discrimination is to retain the greatest number of munitions detections from the anomaly list while rejecting the maximum number of anomalies arising from nonmunitions items. The efficiency measures the fraction of detected munitions retained by the discrimination, while the rejection ratio measures the fraction of false alarms rejected. Both measures are defined relative to the entire response list, i.e., the maximum munitions detectable by the sensor and its accompanying clutter detection rate/false positive rate or background alarm rate.

---

[1]Strictly speaking, ROC curves plot the $P_d$ versus $P_{ba}$ over a predetermined and fixed number of detection opportunities (some of the opportunities are located over munitions and others are located over clutter or blank spots). In an open field scenario, each system suppresses its signal strength reports until some bare-minimum signal response is received by the system. Consequently, the open field ROC curves do not have information from low signal-output locations, and, furthermore, different contractors report their signals over a different set of locations on the ground. These ROC curves are thus not true to the strict definition of ROC curves as defined in textbooks on detection theory. Note, however, that the ROC curves obtained in the blind grid test sites are true ROC curves.

Efficiency (E): $E = P_d^{disc}(t^{disc})/P_d^{res}(t_{min}^{res})$: Measures (at a threshold of interest) the degree to which the maximum theoretical detection performance of the sensor system (as determined by the response stage tmin) is preserved after application of discrimination techniques. Efficiency is a number between 0 and 1. An efficiency of 1 implies that all of the munitions initially detected in the response stage were retained at the specified threshold in the discrimination stage, $t^{disc}$.

False Positive Rejection Rate ($R_{fp}$): $R_{fp} = 1 - [P_{fp}^{disc}(t^{disc})/P_{cd}^{res}(t_{min}^{res})]$: Measures (at a threshold of interest) the degree to which the sensor system's false positive performance is improved over the maximum false positive performance (as determined by the response stage tmin). The rejection rate is a number between 0 and 1. A rejection rate of 1 implies that all emplaced clutter initially detected in the response stage were correctly rejected at the specified threshold in the discrimination stage.

Background Alarm Rejection Rate ($R_{ba}$): $R_{ba} = 1 - [P_{ba}^{disc}(t^{disc})/P_{ba}^{res}(t_{min}^{res})]$. Measures the degree to which the discrimination stage correctly rejects background alarms initially detected in the response stage. The rejection rate is a number between 0 and 1. A rejection rate of 1 implies that all background alarms initially detected in the response stage were rejected at the specified threshold in the discrimination stage.

FISHER’S EXACT TEST

Fisher’s exact test for differences in probabilities (or 2 by 2 contingency table) is used to analyze two samples drawn from two different populations to see if both populations have the same or different proportions of elements in a certain category. More specifically, two random samples are drawn, one from each population, to test the null hypothesis that the probability of some specified event is the same for both populations.

Two-sided tests are completed to compare the performance factors for the capped and uncapped areas. The confidence level for the tests was 0.1. For the generic 2 by 2 contingency table defined in Table A-1, the probability associated with that exact probability of obtaining that table is given below:

$$p = \frac{\binom{a+b}{a}\binom{c+d}{c}}{\binom{n}{a+c}}$$

The p-value for the two-sided test is obtained by summing the probabilities for each possible table given constant margins whose probability is less than or equal to the probability of the given table.

TABLE A-1. GENERIC 2 BY 2
CONTIGENCY TABLE

| | **Capped** | **Uncapped** |
|---|---|---|
| Detected | a | c |
| Not detected | b | d |

An example follows that illustrates Fisher's exact test using fabricated detection results for capped and uncapped areas. It should be noted that a significant result does not prove a cause-and-effect relationship exists between the two populations of interest; however, it does serve as a tool to indicate a change in system performance at a large enough level than can be accounted for merely by chance or random variation. Note also that a result that is not significant indicates that there is not enough evidence to declare that anything more than chance or random variation within the same population is at work between the two data sets being compared.

Demonstrator X achieves the following overall results for $P_d^{res}$ after surveying the Small Munitions Test Site using the same system (results indicate the number of munitions detected divided by the number of munitions emplaced):

| Capped | Uncapped |
|---|---|
| 3/10 = 0.70 | 9/10 = 0.90 |

$P_d^{res}$: CAPPED versus UNCAPPED. Using the example data above to compare probabilities of detection in the response stage, 3 munitions out of 10 emplaced munitions items were detected in the capped area while 9 munitions out of 10 emplaced were detected in the uncapped area. The resulting contingency table is given in Table A-2.

TABLE A-2. EXAMPLE
CONTINCY TABLE

| | **Capped** | **Uncapped** |
|---|---|---|
| Detected | 3 | 9 |
| Not detected | 7 | 1 |

The probability associated with this exact table is 0.00953. Summing all possible tables given constant margins whose probability is less than or equal to 0.00953 results in a p-value of 0.020. Since 0.020 is less than the specified confidence level 0.1, the result of the test is that there is a statistically significant difference between the capped and uncapped areas with respect to the $P_d^{res}$ factor. While a significant result does not prove a cause-and-effect relationship exists, it does indicate that the difference in detection ability between the capped and uncapped areas of Demonstrator X's system is unlikely to be explained completely by chance or random variation.

## APPENDIX B. DAILY WEATHER LOGS

| Date, 2013 | Time, EST | Average Temperature, °F | Total Precipitation, in. |
|---|---|---|---|
| 14 May | 0700 | 42.4 | 0.00 |
| | 0800 | 48.7 | 0.00 |
| | 0900 | 51.4 | 0.00 |
| | 1000 | 53.2 | 0.00 |
| | 1100 | 54.9 | 0.00 |
| | 1200 | 56.5 | 0.00 |
| | 1300 | 58.1 | 0.00 |
| | 1400 | 59.9 | 0.00 |
| | 1500 | 60.4 | 0.00 |
| | 1600 | 60.3 | 0.00 |
| | 1700 | 60.4 | 0.00 |
| 15 May | 0700 | 55.2 | 0.00 |
| | 0800 | 58.8 | 0.00 |
| | 0900 | 62.2 | 0.00 |
| | 1000 | 62.1 | 0.00 |
| | 1100 | 64.4 | 0.00 |
| | 1200 | 64.2 | 0.00 |
| | 1300 | 64.8 | 0.00 |
| | 1400 | 64.8 | 0.00 |
| | 1500 | 68.4 | 0.00 |
| | 1600 | 72.3 | 0.00 |
| | 1700 | 74.3 | 0.00 |
| 16 May | 0700 | 64.6 | 0.00 |
| | 0800 | 68.2 | 0.00 |
| | 0900 | 71.2 | 0.00 |
| | 1000 | 72.9 | 0.00 |
| | 1100 | 74.7 | 0.00 |
| | 1200 | 74.7 | 0.00 |
| | 1300 | 76.3 | 0.00 |
| | 1400 | 76.8 | 0.00 |
| | 1500 | 78.6 | 0.00 |
| | 1600 | 79.7 | 0.00 |
| | 1700 | 79.5 | 0.00 |

EST = Eastern Standard Time.

# APPENDIX C. SOIL MOISTURE

| Date: 14 May 2013 | | | |
|---|---|---|---|
| **Probe Location** | **Layer, in.** | **AM Reading, %** | **PM Reading, %** |
| Small munitions grid | 0 to 6 | 17.8 | 17.7 |
| | 6 to 12 | 28.9 | 28.8 |
| | 12 to 24 | 32.8 | 32.6 |
| | 24 to 36 | 35.6 | 35.8 |
| | 36 to 48 | 46.3 | 46.2 |
| Small munitions calibration grid | 0 to 6 | | |
| | 6 to 12 | 18.7 | 18.6 |
| | 12 to 24 | 27.9 | 27.9 |
| | 24 to 36 | 28.8 | 28.7 |
| | 36 to 48 | 32.5 | 32.6 |
| | | 49.9 | 49.8 |

| Date: 15 May 2013 | | | |
|---|---|---|---|
| **Probe Location** | **Layer, in.** | **A.M. Reading, %** | **P.M. Reading, %** |
| Small munitions grid | 0 to 6 | 17.6 | 17.5 |
| | 6 to 12 | 28.9 | 28.9 |
| | 12 to 24 | 32.4 | 32.3 |
| | 24 to 36 | 35.7 | 35.6 |
| | 36 to 48 | 46.0 | 45.8 |
| Small munitions calibration grid | 0 to 6 | | |
| | 6 to 12 | 18.5 | 18.5 |
| | 12 to 24 | 27.7 | 27.6 |
| | 24 to 36 | 28.6 | 28.5 |
| | 36 to 48 | 32.4 | 32.3 |

| Date: 16 May 13 | | | |
|---|---|---|---|
| **Probe Location** | **Layer, in.** | **A.M. Reading, %** | **P.M. Reading, %** |
| Small munitions grid | 0 to 6 | | - |
| | 6 to 12 | | - |
| | 12 to 24 | | - |
| | 24 to 36 | | - |
| | 36 to 48 | | - |
| Small munitions calibration grid | 0 to 6 | 17.3 | |
| | 6 to 12 | 28.5 | |
| | 12 to 24 | 32.0 | |
| | 24 to 36 | 35.2 | |
| | 36 to 48 | 45.4 | |

# APPENDIX D. DAILY ACTIVITY LOGS

| Date, 2013 | No. of People | Area Tested | Status Start Time, hr | Status Stop Time, hr | Duration min. | Operational Status | Operational Status Comments | Track Method | Pattern | Field Conditions | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 14 May | 3 | Calibration Lanes | 0900 | 1430 | 330 | Initial Setup | Initial Mobilization | GPS | Linear | Sunny | Cool |
| 14 May | 3 | Calibration Lanes | 1430 | 1445 | 15 | Calibration | Calibration | GPS | Linear | Sunny | Cool |
| 14 May | 3 | Calibration Lanes | 1445 | 1455 | 10 | Collecting Data | Collect Data Small Munitions Calibration Grid | GPS | Linear | Sunny | Cool |
| 14 May | 3 | Open Field | 1455 | 1605 | 70 | Collecting Data | Collect Data Small Munitions Test Site | GPS | Linear | Sunny | Cool |
| 14 May | 3 | Open Field | 1605 | 1610 | 5 | Calibration | Calibration | GPS | Linear | Sunny | Cool |
| 14 May | 3 | Open Field | 1610 | 1630 | 20 | Daily Start, Stop | Equipment Breakdown | GPS | Linear | Sunny | Cool |
| 15 May | 3 | Open Field | 1410 | 1520 | 70 | Daily Start, Stop | Set Up Equipment | GPS | Linear | Rainy | Cool |
| 15 May | 3 | Open Field | 1520 | 1550 | 30 | Calibration | Calibration | GPS | Linear | Rainy | Cool |
| 15 May | 3 | Open Field | 1550 | 1610 | 20 | Collecting Data | Collect Data Small Munitions Test Site | GPS | Linear | Rainy | Cool |
| 15 May | 3 | Open Field | 1610 | 1620 | 10 | Weather Issue | Weather Issue Raining | GPS | Linear | Rainy | Cool |
| 15 May | 3 | Open Field | 1620 | 1630 | 10 | Collecting Data | Collect Data Small Munitions Test Site | GPS | Linear | Rainy | Cool |
| 15 May | 3 | Open Field | 0735 | 0800 | 25 | Calibration | Calibration | GPS | Linear | Rainy | Cool |
| 15 May | 3 | Open Field | 0800 | 0855 | 55 | Daily Start, Stop | Set Up Equipment | GPS | Linear | Rainy | Cool |
| 15 May | 3 | Open field | 0855 | 0905 | 10 | Collecting data | Collect Data Small Munitions Test Site | GPS | Linear | Rainy | Cool |
| 15 May | 3 | Open field | 0905 | 1020 | 75 | Calibration | Calibration | GPS | Linear | Rainy | Cool |
| 15 May | 3 | Open field | 1020 | 1035 | 15 | Break/Lunch | Lunch | GPS | Linear | Rainy | Cool |
| 15 May | 3 | Open field | 1035 | 1130 | 55 | Daily Start, Stop | Set Up Equipment | GPS | Linear | Rainy | Cool |
| 15 May | 3 | Open field | 1130 | 1140 | 10 | Calibration | Calibration | GPS | Linear | Rainy | Cool |
| 15 May | 3 | Open field | 1140 | 1245 | 65 | Collecting Data | Collect Data Small Munitions Test Site | GPS | Linear | Rainy | Cool |
| 15 May | 3 | Open field | 1245 | 1340 | 55 | Calibration | Calibration | GPS | Linear | Rainy | Cool |
| 15 May | 3 | Open field | 1340 | 1350 | 10 | Daily Start, Stop | Equipment Breakdown | GPS | Linear | Rainy | Cool |

GPS = Global Positioning System

| Date, 2013 | No. of People | Area Tested | Status Start Time, hr | Status Stop Time, hr | Duration min. | Operational Status | Operational Status Comments | Track Method | Pattern | Field Conditions | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 16 May | 3 | Calibration Lanes | 0810 | 0845 | 35 | Daily Start, Stop | Set Up Equipment | GPS | Linear | Sunny | Cool |
| 16 May | 3 | Calibration Lanes | 0845 | 0855 | 10 | Calibration | Calibration | GPS | Linear | Sunny | Cool |
| 16 May | 3 | Calibration Lanes | 0855 | 0935 | 40 | Collecting Data | Collect Data Small Munitions Calibration Grid +Calibration Grid | GPS | Linear | Sunny | Cool |
| 16 May | 3 | Calibration Lanes | 0935 | 1140 | 125 | Demobilization | Demobilization | GPS | Linear | Sunny | Cool |

## APPENDIX E. REFERENCES

1. Standardized UXO Technology Demonstration Site Handbook, DTC Project No. 8-CO-160-000-473, Report No. ATC-8349, March 2002.

2. Aberdeen Proving Ground Soil Survey Report, October 1998.

3. Data Summary, UXO Standardized Test Site: APG Soils Description, May 2002.

# APPENDIX F. ABBREVIATIONS

| | | |
|---|---|---|
| APG | = | Aberdeen Proving Ground |
| ATC | = | U.S. Army Aberdeen Test Center |
| ATSS | = | Aberdeen Test Support Services |
| BAR | = | background alarm rate |
| DGPS | = | Differential Global Positioning System |
| DMM | = | discarded military munitions |
| EQT | = | Environmental Quality Technology |
| ERDC | = | U.S. Army Corps of Engineers Engineering Research and Development Center |
| EST | = | Eastern Standard Time |
| ESTCP | = | Environmental Security Technology Certification Program |
| GPS | = | Global Positioning System |
| GT | = | ground truth |
| HEAT | = | high-explosive antitank |
| MM | = | military munitions |
| $P_{ba}$ | = | probability of background alarm |
| $P_{cd}$ | = | probability of clutter detection |
| $P_d$ | = | probability of detection |
| $P_{fp}$ | = | probability of false positive |
| POC | = | point of contact |
| QA | = | quality assurance |
| QC | = | quality control |
| $R_{ba}$ | = | background alarm rejection |
| $R_{fp}$ | = | false positive rejection |
| $R_{halo}$ | = | Halo Radius |
| ROC | = | receiver-operating characteristic |
| SERDP | = | Strategic Environmental Research and Development Program |
| TDSS | = | Threat Detection and Systems Survivability |
| TEMTADS | = | TEM |
| TOI | = | targets-of-interest |
| USAEC | = | U.S. Army Environmental Command |
| UXO | = | unexploded ordnance |
| YPG | = | U.S. Army Yuma Proving Ground |

## APPENDIX G. DISTRIBUTION LIST

ATEC PROJECT NO. 2011-DT-ATC-DODSP-F0292

| Addressee | No. of Copies |
|---|---|
| Program Manager<br>SERDP/ESTCP<br>Munitions Management<br>ATTN: Mr. Herb Nelson<br>4800 Mark Center Drive, Suite 17D08<br>Alexandria, VA 22350-3600 | 1 |
| Battelle<br>ATTN: Mr. Jeff Gamey<br>100A Donner Drive<br>Oak Ridge, TN 37830 | 1 |
| Commander<br>U.S. Army Aberdeen Test Center<br>ATTN: TEDT-AT-SLE (Mr. J. Stephen McClung)<br>400 Colleran Road<br>Aberdeen Proving Ground, MD 21005-5059 | 1 |
| Defense Technical Information Center<br>8725 John J. Kingman Road, Suite 0944<br>Fort Belvoir, VA 22060-6218 | PDF |

Secondary distribution is controlled by Program Manager, SERDP/ESTCP, Munitions Management, ATTN: Mr. Herb Nelson.